DOCUMENT RESUME

ED 394 985 TM 024 638

AUTHOR Boshuizen, Henny P. A.; And Others
TITLE Monitoring the Development of Expertise in a Problem-Based Curriculum.
PUB DATE Apr 95
NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Behavioral Sciences; Case Studies; Clinical Diagnosis; Educational Assessment; Evaluation Methods; Foreign Countries; Higher Education; *Knowledge Level; Medical Education; *Medical Students; *Skill Development; *Student Evaluation; Test Construction; *Test Validity
IDENTIFIERS *Expertise; Monitoring; Netherlands; *Problem Based Learning

ABSTRACT

A study was undertaken to investigate the validity of a progress test, the Maastricht Progress Test, that was designed to measure knowledge and clinical reasoning growth in a problem-based medical curriculum. Scores and subscores of about 40 students per year (total sample of 195) on the different categories of the progress test were compared with scores on a clinical reasoning test. Both tests revealed the same pattern of increasing scores over the years, and they had a high correlation. Further analyses revealed that the clinical science component, in particular, of the progress test explained the variations in the clinical reasoning test scores. Knowledge of behavioral sciences had a small but independent contribution. Outcomes are discussed from the perspectives of research and theory on the development of medical expertise, and educational consequences are discussed. An appendix contains an example problem. (Contains one table, three figures, and four references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 394 985

Monitoring the development of expertise in a problem-based curriculum

Henny P.A. Boshuizen, Maureen Machiels-Bongaerts, Henk G. Schmidt, &

Hélène Hermans*

University of Limburg, the Netherlands.

*Antonie van Leeuwenhoek Ziekenhuis, the Netherlands.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Correspondence and requests for reprints to

Dr. Henny P.A. Boshuizen

University of Limburg,

Dept. of Educational Research and Development

PO Box 616

6200 MD Maastricht

Netherlands

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H.P.A. Boshuizen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

MONITORING EXPERTISE DEVELOPMENT

Presented at the Annual Meeting of AERA, San Francisco, April 1995

TM024638

BEST COPY AVAILABLE

Abstract

Purpose of the present study was to investigate the validity of the Progress Test that was especially designed for measuring the knowledge and clinical reasoning growth in a Problem-Based medical curriculum. Scores and subscores of the students on the different categories of the Progress Test were compared with the scores on a Clinical Reasoning Test. Both tests revealed the same pattern of increasing scores over the years - and had a high correlation. Further analyses revealed that especially the Clinical Sciences component in the Progress Test explained the variations in the Clinical Reasoning Test scores. Knowledge of Behavioral Sciences had a small but independent contribution. Knowledge of the Behavioral Sciences did not have this independent effect. These outcomes were discussed from the perspective of development of medical expertise research and theory. Educational consequences are discussed.

Introduction

One of the reasons for promoting problem-based learning is that it is conjectured to encourage self-directed learning in students. In this view on teaching and learning self-directed learning should be preferred over teacher-directed learning. One of the reasons is that students themselves know better than their teachers what they know and do not know and hence what they should give attention to in the ensuing period. Another reason concerns motivation: students should be allowed to pursue questions they are interested in at a specific moment. Intrinsic motivation is thought to be an important determinant for time and effort put into studying and later results. As a consequence, a common feature of problem-based curricula is that students are responsible for their own learning. Teachers or book lists do not prescribe what students have to learn during a specific period; the students themselves decide what they will study. Their decisions are aided by the problems they are working on. While working on such problems (that have been carefully designed by the faculty in order to be able to fulfil this role) they analyze what they know about that issue and what they apparently do not know. The students also do an assessment of the level of detail they want to reach for the moment. Finally,tudents will also decide for themselves which media they want to use. Possibilities are the traditional books, audiovisuals, computer simulations, interview with an expert, field work, etc.

The learning objectives that will be pursued by the individual students will be similar in many respects, but will also differ as a consequence of differences in prior knowledge and interest. This relative freedom of the students has as a consequence that it is very difficult for the faculty to formulate rigorous objectives per course. The final consequence of the view on teaching and learning is that students cannot be passed or failed based on a test that is designed as a traditional end-of-semester test. As an alternative the Medical School of the University of Limburg developed the Maastricht Progress Test. This Progress Test is designed as an exit level test: The cognitive curriculum objectives are translated into true-false items[1]. Together these items should cover all the

[1]Students are allowed to skip those items they have no knowledge of; guessing is not really discouraged, although students sometimes feel that the correction made for guessing is meant for that purpose.

issues a graduate is supposed to know. For each test a sample of 300-400 of such items is taken. An example is:

(given) A patient with renal disorder shows metabolic acidosis.

(question) Hypoventilation contributes to compensating for this acidosis;

~~TRUE~~/FALSE.

Each student, freshman through near-graduate, will sit for the test four times a year. Per test passing scores are calculated for each class. Passing scores set for freshmen are of course much lower than for final year students (Verwijnen, Imbos, Snellen, et al. 1982). Many years of experience with Progress Testing show that the scores of the students continuously increase over the years.

The Progress Test is not only used as an assessment instrument, it also serves as a means of feedback. Students receive detailed reports including their total score, scores on the biomedical, clinical and Behavioral Sciences in general and per subject matter domain. Students also receive reports on individual items and one or more references toward literature that can be checked. On the whole, the Progress Test has shown to be a valuable instrument for assessment and feedback.

A problem with the Progress Test format is that it has fixed response questions. Theoretically, it is possible to use MCQ or true-false questions to assess the students' knowledge about diagnosis, prognosis, etiology, treatment and management of diseases as well as the application of that knowledge on clinical cases. However, genuinely most questions only address the knowledge level. Problem solving and clinical reasoning items are rare; only a few questions per Progress Test can be classified as knowledge application questions pertaining to short cases. Furthermore, fixed response questions with only a few alternatives (two in this case) allow the students to rely on recognition or to reason back from the set alternatives. By doing so, students can circumvent active hypothesis generation and forward search from the information provided in the case. Both strategies may lead to a good answer and students will apply them in case of uncertainty or incomplete knowledge on the subject. However, the use of these strategies does not mirror the way knowledge is applied in actual practice where hypothesis generation plays an important role. Whether this discrepancy between

behavior in actual practice and in test situation is detrimental to the test validity is hard to say, but these are enough reasons for a more authentic investigation of this issue.

In the present article we explore the validity of the Progress Test by comparing student scores on this instrument with scores on a newly developed Clinical Reasoning Test (Schmidt, Machiels-Bongaerts, Hermans et al., 1995). This test consists of 30 case vignettes and has an open-end format. Hence it requires active hypothesis generation as a necessary step toward the Differential Diagnosis that is asked for (see Appendix 1 for an example). The question that is mainly focussed in the present article is whether the same growth patterns can be found in the Progress Test and in the open answer Clinical Reasoning Test. Therefore it will be investigated whether our students' diagnostic problem solving follows the same developmental path as does knowledge measured by the Progress test. Furthermore, we want to investigate how Progress Test scores and Clinical Reasoning scores hang together during the consecutive years.

Method

Data were collected in October 1993. About 40 students per year[2] were invited to participate in the study. Freshmen, who had only started 6 weeks before, were not included, resulting in three preclinical groups: 2nd-, 3rd- and 4th-year students. In the clinical period an extra criterion was used, i.e., the number of clerkships completed. This was done because some students have to accept a waiting time of several months before they can first participate in the clinical rotations. Ro-1 students, who were 5th-year students, had recently started their first clerkship. Ro-2 students were 6th-year students. They had recently completed their third clerkship, which could be either internal medicine, surgery or family medicine (the other two had been done before). These three large clerkships take about 3 months each. Ro-3 students were about to graduate or had graduated recently. Administratively these students are 6th-year students as well; they are only a little bit delayed. In fact, they are about one year ahead of the Ro-2s. Their delay can be due to waiting times, extracurricular

[2]The whole curriculum takes six years, four preclinical and two clinical years. Graduation can take place during the whole year, as soon as the student has fulfilled all requirements.

activities undertaken, etc. Normally, these students do not have better or lower marks than those who graduate before September 1. 223 students participated: 40 second-year students, 41 third-year students, 41 fourth-year students, 40 Ro-1s, 20 Ro-2s, and 41 Ro-3 students.

These students took the Clinical Reasoning Test. This test consisted of 30 vignettes with known diagnosis, covering all organ systems, with the exception of psychiatry. Psychiatry was not included because psychiatric case descriptions are incompatible with the proposed length of the vignettes. Students were asked to read the cases, and to come up with a differential diagnosis. Explanations or justifications were not asked for. The differential diagnoses were scored as follows: if the intended diagnosis was in the first place of the list, 2 points were given; if it was on another place than the first, 1 point was credited. Four cases yielded diagnoses with one or more sub-diagnoses (e.g., case 20 was an acute pancreatitis case with subdiagnoses stone, obstructing bile flow). Students who included such subdiagnoses received some bonus point (maximum of 7). Four cases were excluded because domain experts had the opinion that alternative diagnoses were too plausible.

Of the students who participated in the study, scores on the Progress Tests of September and December 1993 (the first one preceded the period in which data were collected, the second one followed it) were obtained. The students' scores are found by taking the number of items answered correctly, corrected for guessing by subtracting the number of items answered incorrectly. Reliabilities calculated over all students were .979 and .974 , mean reliabilities per year group were .886 and .878[3]. Some students had only taken one test. In that case the score on one test was used. Those students who had not participated in both tests were excluded from the analyses[4]. As a result the samples consisted of 39 (40 originally) second-year students, 41 (41) third-year students, 31 (41) fourth-year students, 39 (40) Ro-1s, 20 (20) Ro-2s, and 25 (41) Ro-3 students. Students received a small financial compensation for their participation.

---

[3]Courtesy to Ron Hoogeboom.

[4]Except for illness, the most common reason for not sitting for the Progress Test is having obligations elsewhere that do not allow travel to Maastricht, e.g., because the student is abroad for electives. Most often this occurs at the end of the fourth year, after a student has finished the preclinical program. Another reason for non-participation is found in the sixth year students group. Students who have taken 24 tests and have obtained enough passing scores are no longer required to sit for the test.

BEST COPY AVAILABLE

Results and Discussion

Insert Figure 1 about here

Figure 1 shows the Progress Test results of the six groups. Groups differ significantly ($F(5, 189) = 77.191$, $p < .0001$); all Newman-Keuls comparisons except those between Ro-2 and Ro-3 are significant. The groups show the same increase over the years that is commonly found on this test. Comparisons of the group means and the population means that are depicted in the same graph (placed within brackets) suggest that the subjects selected for this study do not deviate dramatically from their peers. Differences between sample mean and population mean are never greater than 1.9. Notice that the Ro-2 and -3 groups are compared with the same population value (39.32), the mean score of the 6th-year students.

Insert Figure 2 about here

Figure 2 shows the results of the same students on the Clinical Reasoning Test. Again groups differ significantly ($F(5, 189) = 119.802$, $p < .0001$). The same pattern in the Newman-Keuls comparisons were found. The curves of the Clinical Reasoning Test and the Progress Test have basically the same shape, which is expressed in a correlation of .85 (72% common variance). Calculated at the groups level this correlation drops dramatically, ranging from .30 for the 2nd-year students to .61 for the 3rd-year group. Mean correlation within groups is .49.

Such a discrepancy between correlations in the whole group and in subgroups should be expected as an effect of restriction of range. It might, however, also indicate that although both tests have a large common basis, they measure partly different constructs, i.e., that other factors than pure knowledge (which is measured by the Progress Test) play a role in clinical reasoning at least at different stages of development. In order to investigate this explanation simple regression analyses

and stepwise multiple regression analyses per group and over all students were performed. Variables used were: Clinical Reasoning score, Group (number of years in the curriculum), Progress Test score, and three subscores of the Progress Test: Biomedical, Clinical and Behavioral Sciences. The correlations over all students are shown in Table 1. All correlations are high, except for those with the Behavioral Sciences subscores.

---

insert Table 1 about 1

---

insert Figure 3 about here

A similar picture emerges at the groups level (see Figure 3). Correlations per group are much lower; again the correlations between the Clinical Reasoning score and the Behavioral Science subscore are lowest. Furthermore, correlations are remarkably low in the 2nd-year students group, as compared with the more knowledgeable groups. This might be an effect of lack of relevant knowledge: a real restriction of range. In the later years knowledge seems to have grown enough to explain at least part of the variance in Clinical Reasoning.

Stepwise-regression analyses per group show that per group usually only one factor could be identified that explains the variation in Clinical Reasoning. But this is not always the same factor. In the second-year group the adjusted R^2 is .742, but only one variable (the Biomedical Sciences sub-score of the Progress Test, $F(1,38) = 119.179$; $p < .0001$). In the third-year group the explanatory power of the Progress Test subscores is much higher: adjusted R^2 was .895. In this case differences in the Clinical Sciences score ($F(1,39 = 26.086$) and the Behavioral Sciences knowledge cause this effect ($F1,39) = 7.644$). In the fourth year the picture changes again: adjusted R^2 is .952, resulting from differences in the Biomedical subscore on the Progress Test ($F(1,29) = 20.607$) and differences in Behavioral knowledge ($F(1,29) = 20.298$). During the clinical years again some switches in the factors explaining clinical reasoning occur. In the Ro-1 group the adjusted R^2 is .961 resulting from the variance the Clinical Sciences scores ($F(1,39) = 997.115$). In the Ro-2 group the adjusted R^2

amounts to .965 due to differences in again the Clinical subscore on the Progress Test ($F(1,19) = 551.262$). Finally, in the Ro-3 group we find an adjusted R^2 of .974, due to differences in the Clinical Sciences subscores ($F(1,23) = 34.065$) and in the Behavioral subscores ($F(1,23) = 6.312$).

The changing pattern of percentages of variance explained within groups and the different factors that take this role is not easy to clarify. The fact that usually not more than one factor explains the differences in Clinical Reasoning score might indicate that within one level of experience the three Progress Test subscores hang together so much that if one factor is forced into the stepwise regression analysis, the other two have no unique variation left that can explain the rest of the variance. Maybe that is why either differences in Biomedical knowledge or in Clinical knowledge take the main part of the variance. This observation agrees with our theory that different kinds of knowledge (biomedical, behavioral and clinical) must become integrated in order to be flexibly used in clinical reasoning (Schmidt & Boshuizen, 1993). It is interesting to see that briefly before two important breaks in the students' developmental path (before the start of the clinical period and before graduation) knowledge of the Behavioral Sciences seems to play an independent though small role. Our theory cannot explain why the latter phenomenon would occur at these moments. Since it has not been reported previously we will not try to clarify it now, and rather see if it will be replicated in later studies.

The main aim of the present study was to explore the validity of the Progress Test. Therefore a stepwise regression analysis of all the data using the Total Progress Test score and Group as predictors for Clinical Reasoning was done. The outcome show that both factors have unique, almost equal contributions (adjusted R^2 is .8092, Group: $F(1,192) = 79.0499$, Progress Test: $F(1,192) = 77.4584$), suggesting that the Progress Test and the Clinical Reasoning Test have less in common than is suggested by the high overall correlation. The finding of the two equal components may cast a shadow on the validity of the Progress Test. However, further analysis amends this conclusion. In this analysis again the three subscores in stead of the Total Progress Test score were used. A major part of the variance in the Clinical Reasoning Test outcomes (adjusted $R2 = .956$)) can be thus explained using three factors: Clinical Sciences subscore ($F(1,191) = 123.907$), Group ($F(1,191) =$

18.178) and, again despite the low correlations, the Behavioral Sciences subscore (F (1,191) = 5.539). The Basic Sciences subscore partial correlation was not higher than .062.

This final analysis conveys the impression that the Clinical Reasoning Test heavily draws on the Clinical Sciences component in the Progress Test. However, the mere fact of going through the curriculum and experiencing medical practice in the clinical rotations has itself a unique, but relatively small contribution to the variance in Clinical Reasoning. A still smaller contribution comes from Behavioral Sciences knowledge. Hence we may conclude that knowledge of the Basic Sciences and of the Behavioral Sciences seem to contribute differently to Clinical Reasoning; probably biomedical knowledge is integrated in clinical knowledge (Schmidt & Boshuizen, 1993) while Behavioral Sciences knowledge is less integrated and can hence play a role of its own. The latter fact is in line with the findings by Hobus, Schmidt, Boshuizen and Patel (1987) who found that Behavioral Sciences knowledge was not yet integrated in clinical knowledge of physicians who had recently graduated but was in the clinical knowledge of more experienced physicians.

The educational conclusions that can be drawn from this study largely depends on policy. The outcomes of this study suggest that the Progress Test is a very valuable instrument for monitoring the students' advance. Despite the format applied and the kinds of questions asked, it correlates very well with the Clinical Reasoning Test scores. This specially applies to the Clinical Sciences subscore of the Progress Test, suggesting that the true/false format does not only access pure factual knowledge, but also addresses clinical reasoning. Depending on the school's aim and policy, the kind of questions asked might however be reconsidered. Provided the Medical School accepts the criterion used in this study, the Clinical Reasoning Test, it might also reevaluate the structure of the Progress Test. It might reduce the allotment of Biomedical Science items in the test and hence reduce their influence on the Total Progress Test score, in favor of the Clinical Sciences. Different weighing factors, maybe even differing per year (group) might serve the same purpose.

References

Hobus, P.P.M., Schmidt, H.G., Boshuizen, H.P.A., & Patel, V.L. (1987). Contextual factors in the activation of first hypotheses: expert-novice differences. *Medical Education, 21,* 471-476.

Schmidt, H.G., & Boshuizen, H.P.A. (1993). On acquiring expertise in medicine. *Educational Psychology Review, 5,* 1-17.

Schmidt, Machiels-Bongaerts, Hermans et al., (1995). The development of diagnostic competence: a comparison between a problem-based, an integrated and a conventianal medical curriculm. Paper to be presented at AERA. San Francisco, April 1995

Verwijnen, G.M., Imbos, Tj., Snellen, H.A.M., et al. (1982). The evaluation system of the medical school of Maastricht. *Assessment and evaluation in higher education, 3,* 225 - 244.

Appendix 1

A sixty-five year old lady visits her family physician. She enters your surgery room with red eyes suggesting that she has been crying. She tells you that she worries a lot because she has been losing so much weight. After you have calmed her down, she tells you in a cascade of words that she has lost twenty-five pounds, although she eats well. She is very worried about this state of affairs, sleeps poorly and is restless and agitated. She does not take any drugs. Her family history displays nothing unusual. Upon physical examination you find a sick, restless woman with a sweaty, warm skin. The thyroid gland is diffusely enlarged. Blood pressure 150/89; pulse rate 140/min. irregular and unequal. The legs show pitting edema. The heart is enlarged and a murmur suggesting mitral valve insufficiency is heard. Lab data: T4 300 nmol/l, T3 10 nmol/l, TSH .05 mU/l ECG: atrium fibrillation accompanied by a high ventricle frequency.

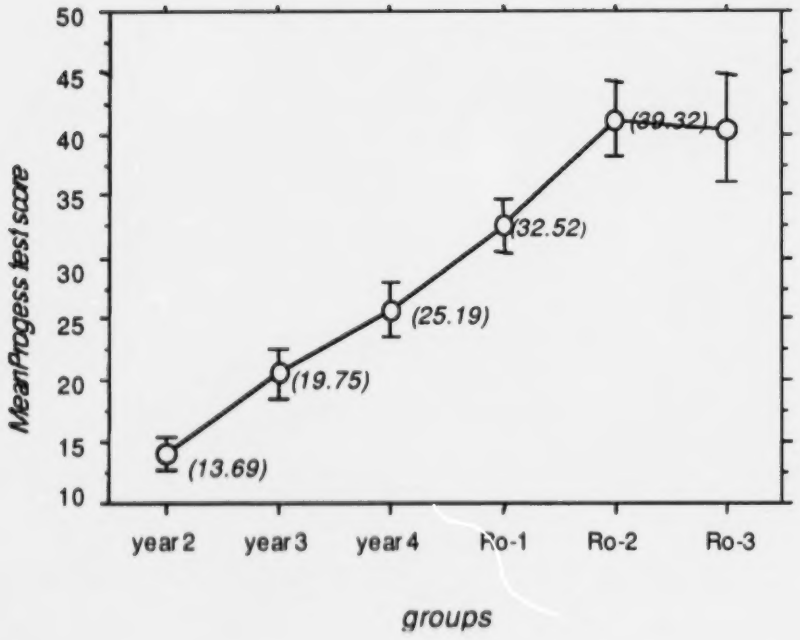

Figure 1: mean percentage score on the two Progress Tests with 95% confidence error-bars Mean scores of the populations are placed between brackets.

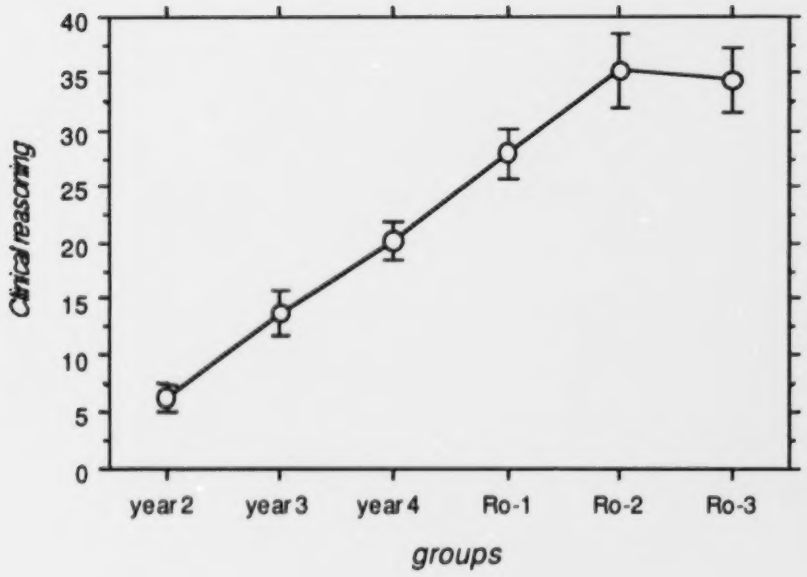

Figure 2: mean clinical reasoning scores for the five groups with 95% confidence error-bars.

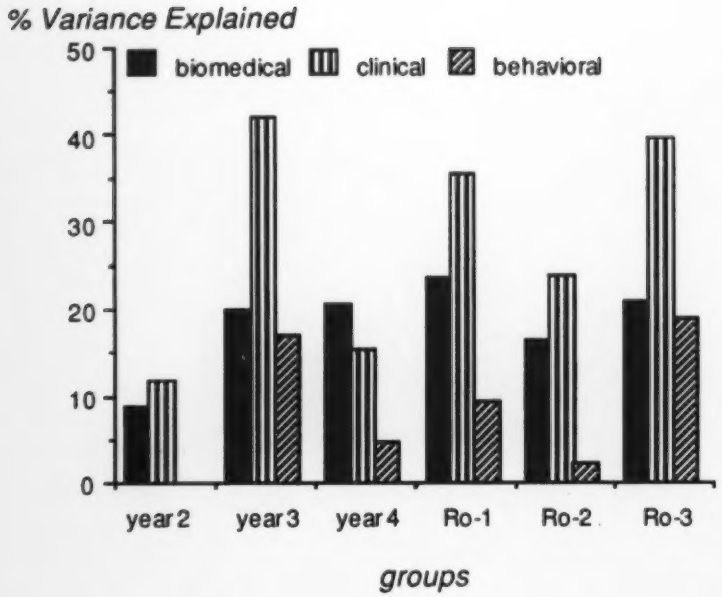

Figure 3. Variance explained (%) in the Clinical Reasoning scores by the Biomedical, Clinical and Behavioral Sciences scores on the Progress Test

Table 1. Correlations between Clinical Reasoning, Group and Progress Test scores (N = 195).

	Clinical Reasoning	Group	Progress Test Total	subscore Biomedical	subscore Clinical	subscore Behavioral
Clinical Reasoning	-					
Group	.8557	-				
Progr Test Total	.8548	.8078	-			
subscore Biomedical	.7690	.7137	.9144	-		
subscore Clinical	.8953	.8899	.9320	.8265	-	
subscore Behavioral	.4315	.3550	.5251	.4127	.3954	-